Repeats in genomic DNA: mining and meaning

Jerzy Jurka

For hundreds of millions of years, perhaps from the very beginning of their evolutionary history, eukaryotic cells have been habitats and junkyards for countless generations of transposable elements, preserved in repetitive DNA sequences. Analysis of these sequences, combined with experimental research, reveals a history of complex 'intracellular ecosystems' of transposable elements that are inseparably associated with genomic evolution.

Addresses
Genetic Information Research Institute, 1170 Morse Avenue,
Sunnyvale, CA 94089, USA; e-mail: jurka@charon.girinst.org

Current Opinion in Structural Biology 1998, 8:333–337

http://biomednet.com/elecref/0959440X00800333

Abbreviations

L1-EN	endonucleolytic domain in L1 reverse transcriptase
LINE	long interspersed nuclear element
LTR	long terminal repeat
MIR	mammalian-wide interspersed repeat
SINE	short interspersed nuclear element
TE	transposable element
TSD	target site duplication

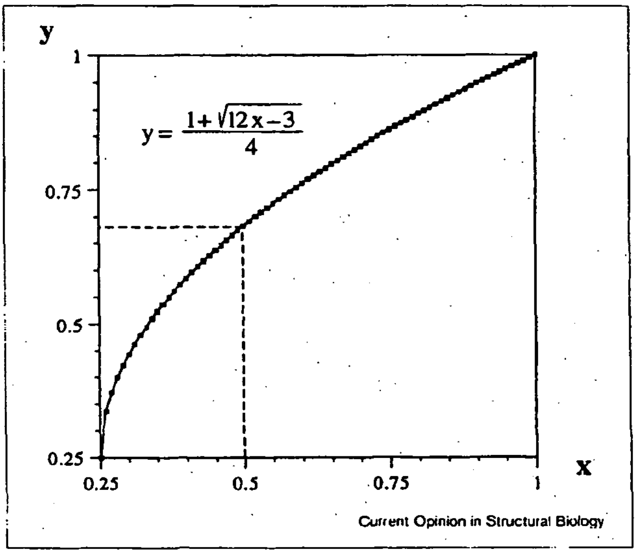
Introduction

Repetitive DNA is a major component of eukaryotic genomes. Understanding its origin, evolution, and genetic impact upon the host DNA is therefore of fundamental importance for genome studies. There are two major groups of repeats in eukaryotic genomes: tandemly repeated satellites, usually confined to specific chromosomal regions; and the repeats interspersed with genomic DNA that are the major focus of this review. Interspersed repeats represent mostly inactive copies of a wide variety of contemporarily and historically active transposable elements (TEs) such as: retroelements and DNA transposons, which can each be further subdivided into distinct classes [1]. Repetitive sequences have been recruited as functional components of eukaryotic genomes, which documents their contribution to genomic evolution [2–6]. They are also an important source of knowledge about the biology of active TEs. The emerging picture, bolstered by recent research, is that TEs are not merely 'parasites'. Rather, they are integral players in genomic evolution, showing either a 'selfish' or an 'altruistic' nature, depending on different evolutionary circumstances.

Reconstruction and analysis of repetitive DNA

As stated above, interspersed repetitive sequences represent inactive (pseudogene) copies of historically or contemporarily active TEs. The study of a new TE usually begins with the identification of its repeated copies, followed by sequence alignment, classification into subfamilies (if applicable) and construction of consensus sequences [7]. Apart from the original TEs themselves, consensus sequences represent the best available approximations of the original active TEs that generated the repeats. Figure 1 illustrates the relationship between the similarities of individual repeats to perfect consensus sequences as compared to similarities between repeats themselves [7]. According to Figure 1, repeats 37–52% similar to each other will be 55–70% similar to their perfect consensus sequences. Without such improvement in similarities, the search for diverse repeats and other biologically meaningful sequence comparisons may be counterproductive.

Figure 1



The similarities between a source gene and its repeats as a function of the similarities between the repeats. The x variable indicates the average similarity between repeats sharing a common source gene; y represents the average similarity of repeats to their source gene that can be approximated by a consensus sequence. For example, repeats that are on average 50% similar to each other will be >68% similar to their ideal consensus sequence. Adapted with permission from [7].

One can reconstruct ancestral TEs even with limited sequence data, especially if individual copies are not very diverse. Additional information may be taken into account, such as the high mutability of CpG dinucleotides or the presence of open reading frames in which nonsense mutations can be reversed. This has been dramatically demonstrated for the *Tc1*-like DNA transposon from fish, named *Sleeping Beauty*, whose transposase was reconstructed from a dozen inactive copies. Its activity has been demonstrated not only in the fish from which it originated, but also in human HeLa cells [8••]. This work, and an earlier study

tabbies
EXHIBIT
F

demonstrating the transfer of a *mariner* element from *Drosophila* to *Leishmania* [9••], are important steps towards application of DNA transposons in genomic studies.

Reconstructions of TEs are very labor intensive and require biological insight but they often remain unpublished. In order to promote the dissemination of this information and to credit the individual effort that goes into producing it, a new electronic publication entitled Repbase Update was established [10•]. Repbase Update represents a systematic attempt to integrate consensus sequence data, nomenclature, biological classification and other relevant information into a coherent resource necessary for sequence studies. To date, over 950 different repetitive sequence families and subfamilies have been compiled from all available eukaryotic sequence data (see Table 1). Of these, over 800 are interspersed repeats. Most interspersed repeats from vertebrates and plants (~80%) have been assigned to one of the following major categories: non-long terminal repeat (LTR) retrotransposons or retroposons also known as SINEs and LINEs, and LTR-retrotransposons including retroviruses and DNA transposons. The remaining nonplant, nonvertebrate repeats come from very diverse species, ranging from protozoans to octopuses, and are temporarily collected under the arbitrary name of 'invertebrates'. In this group, the fraction of interspersed repeats assigned to a particular category is significantly lower (30–40%), mostly due to insufficient comparative sequence data necessary for the construction of reliable consensus sequences. This group of repeats is expected to hold many 'missing links' in our understanding of the origin and evolution of TEs.

Human and rodent sequences can be screened against the most recent version of Repbase Update using public servers [11,12]. Repeat annotation and masking is recommended prior to exon identification [13,14] but Repbase Upgrade is increasingly being used for the direct studies of repetitive DNA.

Table 1

The current content of Repbase Update.

Type of repeats	File name	Number of (sub) families
Human repeats	humrep.ref	284
Alu subfamilies (primate)	humsub.ref	16
Processed pseudogenes (human)	pseudo.ref	20
Rodent repeats	rodrep.ref	157
Other mammalian repeats	mamrep.ref	96
Other vertebrate repeats	vrtrep.ref	74
Plant repeats	plnrep.ref	87
Invertebrate repeats	invrep.ref	222
Simple repeats (microsatellites)	simple.ref	131
Total		**1087**
Unique		**956**

Updated human and rodent collections are also available from public servers for the automatic annotation of DNA sequences [11,12]. Recently computed proportions of repeats in the nonredundant human sequence data are as follows: Alu (12.3%); LINE1 (11.9%); MIR (1.6%); LINE2 (2.1%); LTR retrotransposons and endogenous retroviruses (5.6%); DNA transposons (1.8%); simple repeats (1.4%); other ~0.35%.

The genomic fossil record

The genomic fossil record of past retropositions can be of great value not only for studies of TEs themselves, but also for population and phylogenetic studies of their hosts. For example, young Alu (SINE) subfamilies have been useful for human population studies. To date, there are five known Alu subfamilies (Ya1, Ya5, Yb5, Ya8 and Yb8) actively proliferating in humans [10,15]. Recent innovative studies of 57 Ya5 Alu sequences, 13 of which are polymorphic in the human gene pool, led to an estimate of human effective population size using coalescence theory [16•]. This is only the latest in a series of human population studies based on Alu retroposition.

Turning to older short interspersed nuclear element (SINE) families in mammals, Okada's group [17••] obtained a phylogenetic resolution of the long disputed relationship among whales, ruminants, hippopotamuses and pigs. They have shown that two SINE families, called CHR-1 and CHR-2, are present exclusively in the genomes of whales, ruminants and hippopotamuses, which together form a monophyletic group distinct from that of pigs and camels. This finding contradicts previous phylogenies and illustrates the powerful use of the genomic fossil record in complementing the paleontological record which is particularly difficult to obtain for whales.

Another whale-related development was the identification of homology between the basic units of common satellites and L1 elements, representing the most abundant LINE elements in mammals [18•]. Satellites have long been viewed as a product of unequal crossing over, however, there is no evidence that they can originate *de novo* from nonfunctional 'junk' DNA. The homology between L1 and these satellites supports this scenario and raises many interesting questions about satellite and genomic evolution. Another interesting link between satellites and TEs is the homology between the centromere-associated protein (CENP-B) and the *pogo* family of TEs although biological interpretation of this fact remains tentative [19,20].

Retro (trans) position: a continuation of the transition from the RNA to the DNA world?

Very little is known about the origin of TEs but it is conceivable that the 'TE world', can be traced all the way back to the beginning of the transition from the hypothetical RNA-based genome to the DNA-based one. From this point of view, the entire genomic DNA might have evolved with close participation of TEs, starting with retroposon-like elements. Many TEs might have evolved into parasites, particularly those that can migrate between different hosts, but some may still retain their original properties as 'genome builders'. The examples of *Drosophila* non-LTR retroposons HeT-A and TART, which maintain telomeres in *Drosophila* [21••,22], combined with the recently reported homology between telomerases and reverse transcriptases [23••,24••], bring us closer to this broad perspective [25].

In this context, it may be worthwhile to revisit recent research on the extensively studied mammalian L1 (LINE1) elements. The origin of active mammalian L1 elements remains obscure, but they have produced a succession of numerous subfamilies during the past 100 million years or so [26], and they continue to be active at least in humans and rodents [27•,28]. In spite of their assumed 'selfishness', L1 elements seem to exhibit some remnants of 'altruistic' features that are compatible with active participation in genome evolution. They are responsible for adding over 24% of the DNA to the human genome, only about half of which is L1 DNA (see legend of Table 1 and [12]). Unlike other LINE elements that are parasitized by SINEs homologous to their 3′ ends [29], L1s apparently retropose a large variety of SINE elements and mRNAs ([30••], see below) that have no obvious structural relationship to their own RNA, with the possible exception of poly(A) tails [31]. This is consistent with a recent study demonstrating the ability of L1 reverse transcriptase to efficiently generate cDNA from RNA with no sequence specificity and including transcripts from cellular genes [32•]. Even the affinity of L1 reverse transcriptase for polyadenylated RNA hanging around the ribosomal system [31] may be interpreted as a remnant of the original participation of L1 predecessors in the retroposition of protein encoding RNA. Another relevant property may be the ability of L1 reverse transcriptase to heal chromosomal breaks, although there is some debate as to whether this cannot be attributed to nonhomologous recombination events [33,34].

Diversity and co-evolution of TEs

The genomic fossil record deposited in eukaryotic genomes shows that autonomous TEs tend to be accompanied by nonautonomous companions that are unable to proliferate themselves. Examples include transposon deletion fragments [35,36], SINE elements homologous to 3′ ends of LINE elements [29], and defective LTR retrotransposons, including defective endogenous retroviruses. To multiply, the first group must be able to use transposase from intact DNA transposons, SINE proliferation depends on LINE-encoded reverse transcriptase and the remaining retroelements probably rely on intact viruses for their reproduction. There may be a delicate balance between the autonomous and nonautonomous groups of TEs, analogous to the balance between species in complex ecosystems. Autonomous elements proliferating out of control may destroy their hosts. Nonautonomous elements may destroy themselves by 'successful' competition for the reverse transcriptase or transposase produced by the autonomous TEs. Transposase titration by defective transposons has been discussed among possible factors for the restriction of the activity of mariner-like transposable elements in natural populations [36], although more specialized mechanisms, such as overproduction inhibition, and missense mutation effects are viewed as more prominent events in limiting proliferation of DNA transposons. Multiple LINE1 and SINE (Alu, B1, B2, BC1, etc.) subfamilies in mammals may be viewed as examples of the ongoing co-evolution that is driven by competition for reverse transcriptase [26,30••,37]. LINE2 and mammalian-wide interspersed repeat (MIR) elements [12] might have become extinct as a result of similar competition. Among general mechanisms for the restriction of TEs on the genomic side, suppression by CpG methylation and heterochromatinization have recently been discussed [4,38,39]. Overall, our knowledge of the mechanisms controlling TEs at the genomic level is still fragmentary [40].

Co-evolution between autonomous and nonautonomous elements may not be sufficient to account for the diversity of endogenous retroviruses and retroviral-like elements in mammals. Almost half of all the human repetitive elements deposited in Repbase Update [10•] are either diverse LTRs or fragments of viruses and LTR retrotransposons, although they represent less than 6% of the human genome (see legend of Table 1). In this context, it is worth mentioning a renewed interest in co-evolution between endogenous and exogenous retroviruses that could benefit the host [41,42]. Other related possibilities include recurrent infections and recombinations between distantly related viruses (VV Kapitonov and J Jurka, unpublished data).

Targeting the mammalian genome

Sequence analysis of target site duplications (TSDs) of retroposed elements from mammals [30••], combined with the independent discovery of the endonucleolytic domain in L1 reverse transcriptase (L1-EN, reviewed in [31]), brought about a recent breakthrough in our understanding of retroposon integration in mammals. The consensus sequence of TSDs and adjacent regions for L1, Alu, ID(BC1), B1, B2, and processed pseudogenes is TT|AAAA$(N)_{0-8}$TYTN|R, where R denotes purines, Y represents pyrimidines and N is any base. The vertical bars show predicted positions of breakpoints on the opposite strands of double-stranded DNA [30••,37]. TT|AAAA resembles consensus sequence nicked by the L1-EN [43••], an additional argument implicating L1 reverse transcriptase in the retroposition of nonautonomous retroposons. The general consensus sequence of the TSDs may combine different subclasses of targets. For example, targets beginning with TT|AGAA are longer on average than the targets beginning with TT|AAAA (J Jurka, unpublished data). Different target preferences may be related to different active L1s [27•].

The conserved sequences around both breakpoints in the consensus sequence given above appear to be different from each other, but separate analyses indicate that both sequences are enriched with kinkable TA, CA and TG dinucleotide steps, which suggests a similar mechanism by which both breaks are generated [44•]. This mechanism may be of general significance since the kinkable dinucleotides are conserved in targets both for DNA transposons and for insertion elements in bacteria [44•].

In analogy to the model of intergration of insect R2 non-LTR retroposon [45], the reverse transcription of mammalian retroposons may be primed by the 3′ DNA ends exposed by nicking. Although self-priming of retroposable RNA has been recently demonstrated *in vitro* [46], its role in the retroposition of mammalian retroposons may be marginal if any.

It has long been known that double-stranded breaks stimulate homologous recombination. Therefore, DNA targets exposed to L1-EN nicking acivity may be recombinational hot spots in mammalian genomes. This may have implications for the understanding of at least some of the fragile chromosomal sites involved in the origin of genetic diseases.

Conclusions

The reverse flow of information from RNA to DNA might have had a definite beginning in the history of life, but it has never ended. It remains an integral part of the ongoing genomic evolution in eukaryotic species. It is manifested in active retroposons and in their fossil record as interspersed repetitive DNA. These are the major conclusions emerging from recent progress in the field. Based on these conclusions, the one-dimensional interpretation of TEs as 'parasites' or 'selfish' elements should be transformed into a more balanced view, with their diverse roles comparable to the biological roles of individual species in evolving ecosystems. As the diverse world of TEs continues to emerge with new sequence data, TEs are increasingly being explored in a broad range of biological problems, from phylogenetic and population studies to genomic engineering.

Acknowledgements

Many outstanding and relevant contributions prior to 1997 could not be reviewed here. I selected a number of broad recent reviews to compensate for this deficiency. I would like to thank Vladimir Kapitonov, Paul Klonowski, Dorothy Munro and Jolanta Walichiewicz for help with editing this manuscript. This work was supported by the National Institutes of Health grant 1 P41 LM06252.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Capy P: **Classification of transposable elements.** In *Molecular Biology Intelligence Unit: Dynamics and Evolution of Transposable Elements.* Edited by Capy P, Bazin C, Hiquet D, Langin T. Georgetown, Texas: Landes Bioscience; 1998:37-52.

2. Brosius J, Tiedge H: **Reverse transcriptase: mediator of genomic plasticity.** *Virus Genes* 1996, **11**:163-179.

3. Levin HL: **It's prime time for reverse transcriptase.** *Cell* 1997, **88**:5-8.

4. Kidwell MG, Lisch D: **Transposable elements as sources of variation in animals and plants.** *Proc Natl Acad Sci USA* 1997, **94**:7704-7711.

5. Tomilin NV: **Control of genes by mammalian retroposons.** *Int Rev Cytol* 1998, in press.

6. Chu WM, Ballard R, Carpick BW, Williams BR, Schmid CW: **Potential Alu function: regulation of the activity of double-stranded RNA-activated kinase PKR.** *Mol Cell Biol* 1998, **18**:58-68.

7. Jurka J: **Approaches to identification and analysis of interspersed repetitive DNA sequences.** In *Automated DNA sequencing and analysis.* Edited by Adams MD, Fields C, Venter JC. San Diego: Academic Press Incorporated; 1994:294-298.

8. Ivics Z, Hackett PB, Plasterk RH, Izsvak Z: **Molecular reconstruction of *Sleeping Beauty*, a *Tc1*-like transposon from fish, and its transposition in human cells.** *Cell* 1997, **91**:501-510.
•• This important work is about the reconstruction of an active transposase from 12 pseudogenes found in eight different fish species and using a modified consensus sequence. The approach used has implications for the reconstruction of other proteins involved in proliferation of transposable elements, for the engineering of new transposable elements, and for genome studies.

9. Gueiros-Filho FJ, Beverley SM: **Trans-kingdom transposition of the *Drosophila* element mariner within the protozoan *Leishmania*.** Science 1997, **276**:1716-1719.
•• The authors demonstrate the efficient transfer of the *Drosophila mauritania* mariner element into the human parasite *Leishmania major*. This, and recent experiments with a reconstructed transposase [8••], clearly demonstrate the feasibility of genetic studies on a wide variety of species using DNA transposable elements.

10. *Repbase Update* 1997 on World Wide Web URL:
• http://www.girinst.org/~server/repbase.html
This is a collective attempt to organize the explosively growing number and variety of repetitive sequences. Repbase Update includes many consensus sequences of transposable elements and their biological characterization that are unreported anywhere else.

11. Genetic Information Research Institute on the World Wide Web URL: http://charon.girinst.org

12. Smit AFA: **The origin of interspersed repeats in the human genome.** *Curr Opin Genet Dev* 1996, **6**:743-748.

13. Burge C, Karlin S: **Prediction of complete gene structures in human genomic DNA.** *J Mol Biol* 1997, **268**:78-94.

14. Claverie JM: **Computational methods for the identification of genes in vertebrate genomic sequences.** *Hum Mol Genet* 1997, **6**:1735-1744.

15. Mighell AJ, Markham AF, Robinson PA: **Alu sequences.** *FEBS Lett* 1997, **417**:1-5.

16. Sherry ST, Harpending HC, Batzer MA, Stoneking M: **Alu evolution in human populations: using the coalescent to estimate effective population size.** *Genetics* 1997, **147**:1977-1982.
• This paper demonstrates a very interesting application of Alu polymorphism for estimating human effective population size during the last 1–2 million years.

17. Shimamura M, Yasue H, Ohshima K, Abe H, Kato H, Kishiro T, Goto M, Munechika I, Okada N: **Molecular evidence from retroposons that whales form a clade within even-toed ungulates.** *Nature* 1997, **388**:666-670.
•• This paper addresses an important phylogenetic problem by innovative exploitation of selected repetitive sequences. This is a powerful example of how the genomic fossil record for some species can be more informative than the paleontological record.

18. Kapitonov V, Holmquist G, Jurka J: **L1 repeat is a basic unit of heterochromatin satellites in cetaceans.** *Mol Biol Evol* 1998, **15**:611-612.
• This work has important implications for the understanding of the origin and evolution of satellite DNA.

19. Halverson D, Baum M, Stryker J, Carbon J, Clarke L: **A centromere DNA-binding protein from fission yeast affects chromosome segregation and has homology to human CENP-B.** *J Cell Biol* 1997, **136**:487-500.

20. Kipling D, Warburton PE: **Centromeres, CENP-B and Tiggerr too.** *Trends Genet* 1997, **13**:141-145.

21. Danilevskaya ON, Arkhipova IR, Traverse KL, Pardue ML: **Promoting in tandem: the promoter for telomere transposon HeT-A and implications for the evolution of retroviral LTRs.** *Cell* 1997, **88**:647-655.
•• This work shows that promoter activity in the retroposan HeT-A is located at its 3′ end, in contrast to other retroposons. Tandemly arranged HeT-A elements share these 3′ promoters with their downstream neighbors. The authors conclude that, because of its unusual structure, HeT-A resembles an evolutionary intermediate between non-LTR and LTR retrotransposons.

22. Pardue ML, Danilevskaya ON, Traverse KL, Lowenhaupt K: **Evolutionary links between telomeres and transposable elements.** *Genetica* 1997, **100**:73-84.

23. Ligner J, Hughes TR, Shevchenko A, Mann M, Lundblad V, Cech TR: **Reverse transcriptase motifs in the catalytic subunit of telomerase.** *Science* 1997, **276**:561-567.
••

Telomerase catalytic subunits were first identified in *Euplotes aediculatus* and *Saccharomyces cerevisiae*, and were shown to contain reverse transcriptase motifs. This paper further demonstrates the fact that the reverse transcriptase motif is essential for normal chromosome telomere replication. This work brings together retroposition and chromosome maintenance and has profound evolutionary implications.

24. •• Nakamura TM, Gregg BM, Chapman KB, Weinrich SL, Andrews WH, Lingner J, Harley CB, Cech TR: **Telomerase catalytic subunit homologs from fission yeast and human.** *Science* 1997, 277:955-959.

This paper reveals that the catalytic subunits of telomerases [23••] have conserved domains common to all reverse transcriptases. These domains also revealed distinct hallmarks and the authors conclude that they represent a deep branch in the evolution of reverse transcriptases, and perhaps originated with the first eukaryote.

25. Eickbush TH: **Telomerase and retrotransposons: which came first?** *Science* 1997, 277:911-912.

26. Smit AFA, Toth G, Riggs AD, Jurka J: **Ancestral mammalian-wide subfamilies of LINE-1 repetitive sequences.** *J Mol Biol* 1995, 246:401-417.

27. • Sassaman DM, Dombroski BA, Moran JV, Kimberland ML, Naas TP, DeBerardinis RJ, Gabriel A, Swergold GD, Kazazian HH Jr: **Many human L1 elements are capable of retrotransposition.** *Nat Genet* 1997, 16:37-43.

This paper estimates the number of active L1 copies in the human genome. Different L1s may account for the presence of different targets for retroposon integration, as discussed in the review.

28. Naas TP, DeBerardinis RJ, Moran JV, Ostertag EM, Kingsmore SF, Seldin MF, Hayashizaki Y, Martin SL, Kazazian HH Jr: **An actively retrotransposing, novel subfamily of mouse L1 elements.** *EMBO J* 1998, 17:590-597.

29. Okada N, Hamada M, Ogiwara I, Ohshima K: **SINEs and LINEs share common 3′ sequences: a review.** *Gene* 1997, 205:229-243.

30. •• Jurka J: **Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons.** *Proc Natl Acad Sci USA* 1997, 94:1872-1877.

This paper shows for the first time that the integration of SINE, L1 and processed retropseudogenes occurs at nonrandom, consensus-defined sequence targets. This strongly links the L1 retroposition machinery to the proliferation of non-LINE retroposons and has implications for understanding of the mechanism of retroposition.

31. Boeke JD: **LINEs and Alus — the polyA connection.** *Nat Genet* 1997, 16:6-7.

32. • Dhellin O, Maestre J, Heidmann T: **Functional differences between the human LINE retrotransposon and retroviral reverse transcriptases for *in vivo* mRNA reverse transcription.** *EMBO J* 1997, 16:6590-6602.

This paper demonstrates the specific and high efficiency of L1 reverse transcription of RNA that has no sequence specificity. This is compatible with 'unselfish' aspects of L1 previously discussed in this review.

33. Teng SC, Kim B, Gabriel A: **Retrotransposon reverse-transcriptase-mediated repair of chromosomal breaks.** *Nature* 1996, 383:641-644.

34. Lauermann V: **DNA repair by recycling reverse transcripts.** *Nature* 1997, 386:31-32.

35. Vos JC, De Baere I, Plasterk RHA: **Transposase is the only nematode protein required for *in vitro* transposition of Tc1.** *Genes Dev* 1996, 10:755-761.

36. Hartl DL, Lozovskaya ER, Nurminsky DI, Lohe AR: **What restricts the activity of mariner-like transposable elements?** *Trends Genet* 1997, 13:197-201.

37. Jurka J, Klonowski P: **Integration of retroposable elements in mammals: selection of target sites.** *J Mol Evol* 1996, 43:685-689.

38. Yoder JA, Walsh CP, Bestor TH: **Cytosine methylation and the ecology of intragenomic parasites.** *Trends Genet* 1997, 13:335-340.

39. Bird A: **Does DNA methylation control transposition of selfish elements in the germline?** *Trends Genet* 1997, 13:469-470.

40. Labrador M, Corces VG: **Transposable element-host interactions: regulation of insertion and excision.** *Annu Rev Genet* 1997, 31:381-404.

41. Van der Kuyl AC: **Endogenous retrovirus sequences and their usefulness to the host.** *Trends Microbiol* 1997, 5:339.

42. Best S, Le Tissier PR, Stoye JP: **Endogenous retroviruses and the evolution of resistance to retroviral infection.** *Trends Microbiol* 1997, 5:313-318.

43. •• Feng Q, Moran JV, Kazazian HH Jr, Boeke JD: **Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition.** *Cell* 1996, 87:905-916.

This breakthrough paper demonstrates the presence of an endonucleolytic domain in L1-encoded reverse transcriptase, implying that reverse transcription in mammals is primed by the 3′ DNA ends that are exposed by nicking, as previously established in insects [45].

44. • Jurka J, Klonowski P, Trifonov EN: **Mammalian retroposons integrate at kinkable DNA sites.** *J Biomol Struct Dyn* 1998, 15:717-721.

Sequence data indicate that the integration of retroposons and other TEs may be associated with the formation of DNA kinks. This suggests the presence of universal structural features associated with the integration of TEs.

45. Luan DD, Korman MH, Jakubczak JL, Eickbush TH: **Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition.** *Cell* 1993, 72:595-605.

46. Shen MR, Brosius J, Deininger PL: **BC1 RNA, the transcript from a master gene for ID element amplification, is able to prime its own reverse transcription.** *Nucleic Acids Res* 1997, 25:1641-1648.